



# Identification of potential Mpro inhibitors for the treatment of COVID-19 by using systematic virtual screening approach

Ashish M. Kanhed<sup>1</sup> · Dushyant V. Patel<sup>2</sup> · Divya M. Teli<sup>3</sup> · Nirav R. Patel<sup>2</sup> · Mahesh T. Chhabria<sup>3</sup> · Mange Ram Yadav<sup>4</sup>

Received: 5 May 2020 / Accepted: 23 July 2020 / Published online: 31 July 2020  
© Springer Nature Switzerland AG 2020

## Abstract

The Corona virus Disease (COVID-19) is caused because of novel coronavirus (SARS-CoV-2) pathogen detected in China for the first time, and from there it spread across the globe creating a worldwide pandemic of severe respiratory complications. The virus requires structural and non-structural proteins for its multiplication that are produced from polyproteins obtained by translation of its genomic RNA. These polyproteins are converted into structural and non-structural proteins mainly by the main protease (Mpro). A systematic screening of a drug library (having drugs and diagnostic agents which are approved by FDA or other world authorities) and the Asinex BioDesign library was carried out using pharmacophore and sequential conformational precision level filters using the Schrodinger Suite. From the screening of approved drug library, three antiviral agents ritonavir, nelfinavir and saquinavir were predicted to be the most potent Mpro inhibitors. Apart from these pralmorelin, iodixanol and iotrolan were also identified from the systematic screening. As iodixanol and iotrolan carry some limitations, structural modifications in them could lead to stable and safer antiviral agents. Screenings of Asinex BioDesign library resulted in 20 molecules exhibiting promising interactions with the target protein Mpro. They can broadly be categorized into four classes based on the nature of the scaffold, viz. disubstituted pyrazoles, cyclic amides, pyrrolidine-based compounds and miscellaneous derivatives. These could be used as potential molecules or hits for further drug development to obtain clinically useful therapeutic agents for the treatment of COVID-19.

---

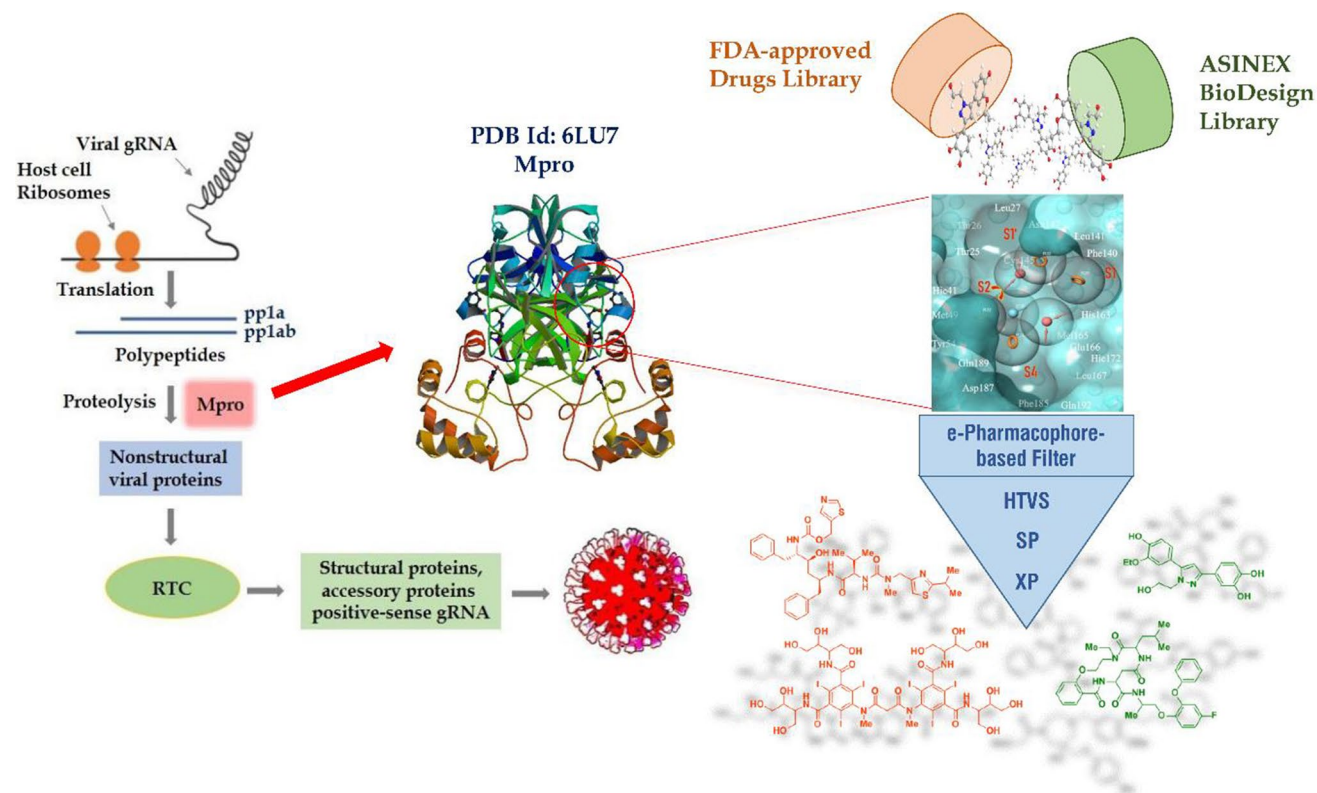
**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11030-020-10130-1>) contains supplementary material, which is available to authorized users.

---

✉ Mange Ram Yadav  
mryadav11@yahoo.co.in

- <sup>1</sup> Shobhaben Pratapbhai Patel - School of Pharmacy and Technology Management, SVKM's NMIMS University, Vile Parle, Mumbai 400056, India
- <sup>2</sup> Faculty of Pharmacy, Kalabhavan Campus, The Maharaja Sayajirao University of Baroda, Vadodara, Gujarat 390001, India
- <sup>3</sup> Department of Pharmaceutical Chemistry, L. M. College of Pharmacy, Navrangpura, Ahmedabad, Gujarat 380009, India
- <sup>4</sup> Parul University, Limda, Waghodia Road, Vadodara, Gujarat, India

## Graphic abstract



**Keywords** Coronavirus · COVID-19 · Severe acute respiratory syndrome · Mpro · Virtual screening · Asinex BioDesign

### Abbreviations

ACE2	Angiotensin-converting enzyme 2
COVID-19	Corona virus disease
CoVs	Coronaviruses
3CLpro	3-Chymotrypsin-like protease
XP	Extra-precision
gRNA	Genomic RNA
HTVS	High-throughput virtual screening
Mpro	Main protease
MERS	Middle east respiratory syndrome
2019-nCoV	Novel coronavirus
PLpro	Papain-like protease
RdRP	RNA-dependent RNA polymerase
RTC	Replicase transcriptase complex
SARS	Severe acute respiratory syndrome
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
sgRNA	Subgenomic RNA
S	Spike
SP	Standard precision
TMPRSS2	Type 2 transmembrane serine protease
WHO	World Health Organization

$\alpha$ -CoV	$\alpha$ -Coronavirus
$\beta$ -CoV	$\beta$ -Coronavirus
$\gamma$ -CoV	$\gamma$ -Coronavirus
$\delta$ -CoV	$\delta$ -Coronavirus

### Introduction

Coronaviruses (CoVs) induced three major outbreaks of respiratory distress syndrome in the last decades including severe acute respiratory syndrome (SARS) in 2003 with an epicenter in Guangdong, China [1], Middle East respiratory syndrome (MERS) in 2012 in Saudi Arabia [2], and novel coronavirus disease (COVID-19) firstly observed in Wuhan province, China, in late 2019 [3]. The COVID-19 symptoms vary from mild to severe that can eventually lead to death. The symptoms usually develop after 2–14 days following the viral exposure which includes fever, cough, shortness of breath and pneumonia. Severe cases showed respiratory, gastrointestinal, hepatic and neurological complications that can lead to death. The human-to-human transmission

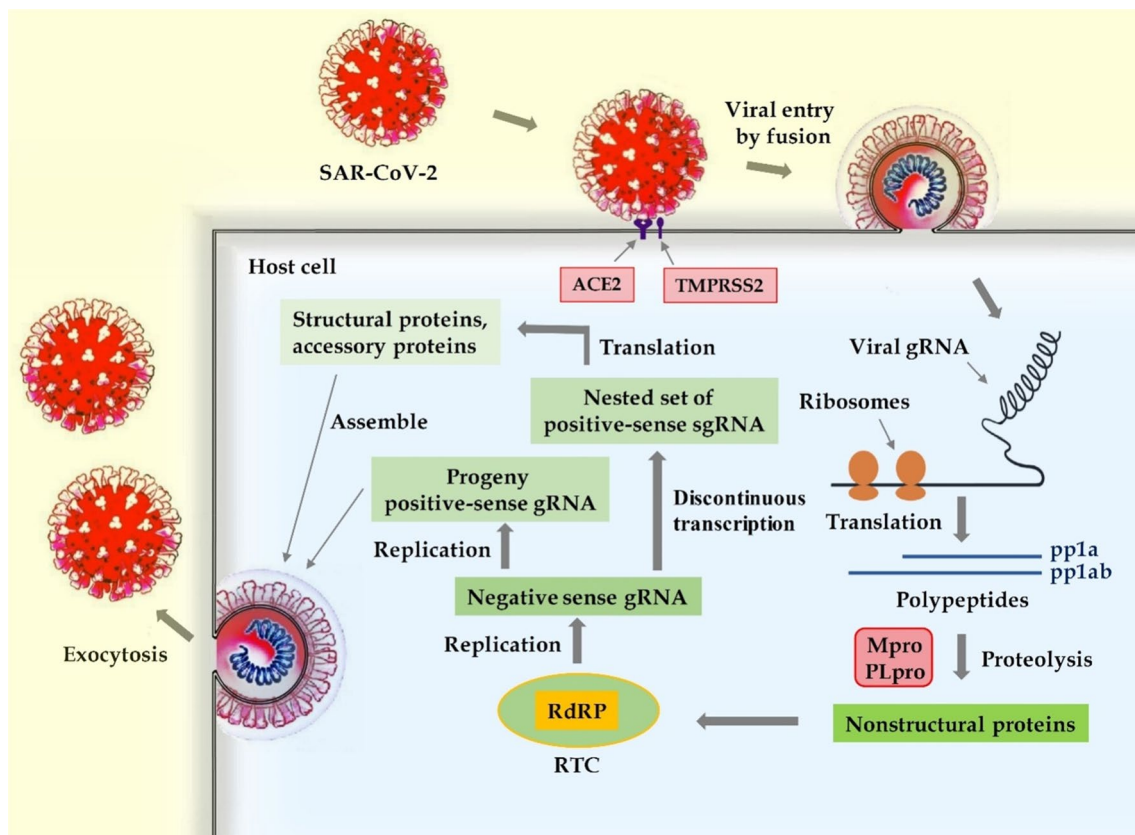
of COVID-19 is reported to occur by respiratory droplets or direct contact with the patients [4, 5].

World Health Organization (WHO) announced COVID-19 as a pandemic on 11th March 2020 [6]. The number of infected cases and deaths by COVID-19 is sharply rising daily. According to the WHO report, as of 18th April 2020, COVID-19 has spread to more than 200 countries and territories with over 2,160,207 confirmed cases and over 146,088 confirmed deaths worldwide. The global fatality rate is presently around 6.76% (calculated as deaths per confirmed cases). More and more number of positive cases has been identified in the USA followed by Spain, Italy, Germany, France, UK, China and Iran [7].

CoVs are a group of genotypically and phenotypically diverse, enveloped and positive-sense viruses carrying single-stranded RNA. They belong to the *Coronaviridae* family that induces respiratory, enteric, hepatic and neurological complications of varying severity in a wide range of animal species and humans [8]. CoVs are divided into four types:  $\alpha$ -coronavirus ( $\alpha$ -CoV),  $\beta$ -coronavirus ( $\beta$ -CoV),  $\gamma$ -coronavirus ( $\gamma$ -CoV) and  $\delta$ -coronavirus ( $\delta$ -CoV). Novel coronavirus (2019-nCoV) also known as severe acute

respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of COVID-19 [9]. SARS-CoV-2, like SARS-CoV and MERS-CoV, belongs to the category of  $\beta$ -coronavirus. Although it is considered to be introduced from bats, the specific source of SARS-CoV-2, animal reservoir and enzootic patterns of transmission remain unresolved.

The lifecycle of SARS-CoV-2 in host cells is briefly discussed here (Fig. 1). The SARS-CoV-2 virus targets the host cells through the viral spike (S) protein. This S protein binds to the ACE2 receptor of the host cells. After the S protein-ACE2 binding, the virus utilizes the host cell receptors (TMPRSS2) and enters into host cell cytosol. After uncoating, the viral gRNA is released into the cytoplasm. Once inside the host cell, viral polypeptides are synthesized using the host cell protein synthesis machinery, which are further processed by viral proteases and the products transferred to RTC. The virus then synthesizes RNA using its RdRP. Viral structural proteins and assembly proteins are also synthesized leading to the completion of assembly and release of progeny viral particles by exocytosis [10, 11]. Generally,  $\beta$ -CoVs produce pp1a (450–500 kDa) and pp1ab



**Fig. 1** Schematic representation of SARS-CoV-2 life cycle in host cells. *gRNA* genomic RNA, *ACE2* angiotensin-converting enzyme 2, *TMPRSS2* type 2 transmembrane serine protease, *RTC* replicase tran-

scriptase complex, *RdRP* RNA-dependent RNA polymerase, *sgRNA* subgenomic RNA

(750–800 kDa) by translation of the genomic RNA [12]. These polypeptides are proteolytically processed into structural and non-structural proteins which are important for the replication and assembly of the virus. This proteolytic cleavage is mediated by the main protease (Mpro) also known as 3-chymotrypsin-like protease (3CLpro) and by papain-like protease (PLpro). Mpro is homodimeric cysteine protease which cleaves the polyprotein at 11 distinct sites generating vital proteins for the viral life cycle. This functional prominence of Mpro in the viral life cycle, along with the lack of closely associated homologues in humans, makes the Mpro a promising target for COVID-19 antiviral drug design [13–17].

The conventional drug discovery approaches offering new drugs into the market demands a long period of time, resources and a huge amount of money with a low success rate. Drug repurposing is an approach of screening of libraries of FDA-approved drugs to identify drug molecules that could be effective for a particular disease. Repurposing FDA-approved drugs hastens the drug development processes, as most of the safety concerns, preclinical testing and formulation development have already been performed for these drugs. Many research groups have carried out the repurposing process, and to date many approved drugs are being tried in more than 100 clinical trial studies, but as of now, none of them turned out to be completely effective for COVID-19 treatment. Unfortunately, no vaccine has been developed/approved yet to combat COVID-19 [18]. To meet this urgent medical need for an effective drug for COVID-19, we planned to identify potential SARS-CoV-2 Mpro inhibitors as possible anti-COVID-19 agents, by carrying out screening of a drug library containing drugs and diagnostic agents (approved by FDA or other world authorities) [19, 20] and the Asinex BioDesign library [21]. In the present report, systematic screenings of these two libraries were carried out using pharmacophore and sequential conformational-precision level filters using the Schrodinger Suite. The resulted scaffolds, mainly from Asinex BioDesign library, are discussed in detail that could be utilized as potential hits for further drug development to find out an acceptable drug treatment of COVID-19.

## Results and discussion

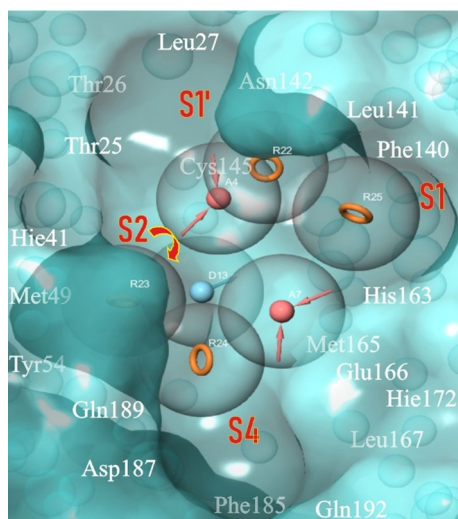
To accomplish the goal, the protein structure of cysteine protease Mpro was obtained from RCSB (PDB code: 6LU7) [14] and the covalent bond between the co-crystallized ligand N3 and the protein catalytic dyad Cys145 was cleaved. After cleavage, Cys145 and the ligand were reconstructed by making necessary changes and the resulted ligand–protein complex was prepared and refined using Protein Preparation Wizard [22] in Schrodinger. This cysteine

protease Mpro having 306 amino acids chain mainly consists of three domains. Domain 1 is from residue 8–101, domain 2 is from 102 to 184, and the domain 3 is from 201 to 303 amino acid sequence connected to domain 2 by a loop having 185–200 amino acid residues. The substrate-binding site on this viral protein is present in a cleft between domains 1 and 2 with a Cys145-His41 catalytic dyad. The major active subsites in the active site of Mpro, where the substrate binds, are defined. The S1 subsite is comprised of Phe140, Leu141, Asn142, His163, Glu166 and His172 amino acids. A small part of S1 subsite is further separated by Asn142 as S1' which is comprised of Thr25, Thr26 and Leu27. Hydrophobic S2 subsite is made up of His41, Met49, Tyr54, Met165 and Asp187. The S4 binding subsite involves Met165, Leu167, Phe185, Gln189 and Gln192 amino acids [14]. Receptor grid was generated on the active site of Mpro protein within Schrodinger Suite. Glide module of Schrodinger was used to carry out the docking study at various precision levels, which include high-throughput virtual screening (HTVS, low precision level), standard precision (SP) and extra-precision (XP). Depending on the precision level applied, Glide internally generates multiple conformations of the ligands under study using ConfGen package. To validate the generated grid, the co-crystallized ligand (N3) was first knocked out, reconstructed and re-docked into the active site of the receptor using the generated grid. Here, the N3 molecule showed a similar pattern of orientation and interactions such as hydrogen bonding with Glu166, Gln189 and Thr190 residues of the active site. The XP docking score between N3 and Mpro protein was  $-7.85$  kcal/mol. This cognate docking was analyzed further by identifying the all-atom RMSD value of the re-docked N3 ligand vis-a-vis the co-crystallized ligand, and it was found to be  $0.452$  Å, which validated the docking protocol.

To identify the possible inhibitors from both the libraries selected for this study, initially, a pharmacophore-based filter was applied. Pharmacophore is the ensemble of steric and electronic features that are necessary to ensure the best possible interactions with a biological target active site. The generation of ligand-receptor/protein complex-based pharmacophore specifies only the important steric and electrostatic features present in that particular ligand, which is co-crystallized with the protein structure. Thus the development of a ligand-receptor/protein complex-based pharmacophore model carries the highest chances of missing out other possible sites for interactions present in the protein active site. Also, to generate a ligand-based pharmacophore model, sufficient number of ligands, with some scaffold similarity, acting against the same target is required which are not available for the target under study to date. Thus, we generated a receptor active site-based pharmacophore model which considers all possible interaction sites in the binding pocket. The inbuilt fragment library was allowed to dock in the grid

generated on the protein active site which checked the interactions of fragments with the residues present in all subsites of the active site, and based on extra-precision (XP) docking scoring function, seven feature-based best pharmacophore hypothesis (AADRRRR) was obtained (Fig. 2). One of the acceptor features was observed in the proximity of His41 with hydrogen on the epsilon nitrogen and Cys145. Another acceptor feature was located near Glu166. The donor feature was observed in close proximity to the carbonyl group of His164. One aromatic feature was observed in the S1 subsite region, whereas one was observed protruding to S1' subsite, one in the S2 subsite region and one in the S4 subsite region of the protein active site explaining these features as either aromatic or hydrophobic in nature.

The resulted hypothesis was checked by aligning the N3 ligand on the pharmacophore hypothesis. Overall, the N3



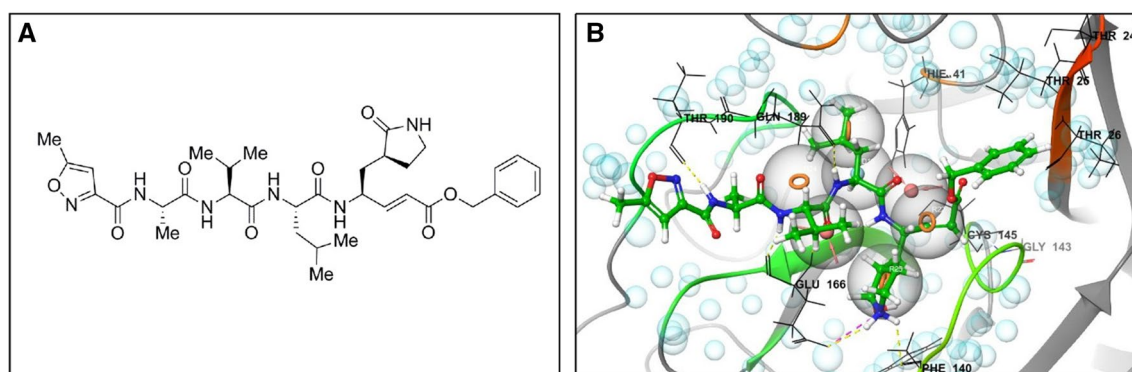
**Fig. 2** Seven feature pharmacophore hypothesis (AADRRRR). (A) Hydrogen bond acceptors, pink spheres with arrows; (D) hydrogen bond donor, sky blue sphere with arrows; (R) ring aromatics, orange torus rings

ligand matched with four pharmacophore features (Fig. 3). One of the acceptor features was observed to be occupied by carbonyl oxygen of N3 which formed a hydrogen bond with NH of Glu166. One of the hydrophobic/aromatic features was occupied by the isobutyl chain of N3 ligand. The (2-oxopyrrolidin-3-yl)methyl moiety of N3 was observed to occupy another hydrophobic/aromatic feature. The but-2-enyl hydrophobic part of N3 ligand inhabited one hydrophobic/aromatic feature. This same research publication [14] has reported six other ligands acting on Mpro. So, these molecules were also assessed on the developed pharmacophore model and were found to match very promisingly with the developed model. These observations validated the developed pharmacophore model. (Data for the other six ligands are shown in the supporting information.)

Using this hypothesis as the first screening filter, the systematic screening of libraries selected under this study was carried out and the finally obtained molecules are discussed here.

The approved drugs and diagnostic agents library was first screened using the developed pharmacophore model followed by the docking protocols using various precision levels (mentioned in experimental section). In the systematic virtual screening, three antiviral drugs were obtained from the approved drug library. All these three drugs ritonavir (**1**), nelfinavir (**2**) and saquinavir (**3**) are viral protease inhibitors. Out of these, ritonavir (**1**) is already in phase II, III and IV clinical trials in combination with some other drug candidates (Phase II: NCT04330690, NCT04276688, NCT04321993, NCT04330586, NCT04332094; Phase III: NCT04261270, NCT04321174, NCT04334928; Phase IV: NCT04291729, NCT04255017, NCT02735707, NCT04286503) [23].

The (thiazoly-5-yl)methyl carbamate oxygen of ritonavir (**1**) was observed engaged in forming hydrogen bonds with Gly143 and Cys145, whereas the thiazolyl ring was observed aligned with polar Thr25, Thr26 and nonpolar Leu27 amino acids of S1' subsite. The phenyl ring on the 1st carbon of



**Fig. 3** (A) 2D structure of N3 ligand; (B) alignment of N3 ligand with the developed pharmacophore model

hexane was observed to be stabilized in the S1 binding pocket, and the phenyl ring present on the 6th carbon of hexane was observed in the S2 binding pocket and made  $\pi$ - $\pi$  interactions with His41. Apart from these interactions, the ligand-receptor/protein complex was further stabilized by hydrogen bonding between the carbonyl oxygen of the butanoyl amino group with Glu166 and NH of the same butanoyl amino group with Gln189 (Fig. 4A).

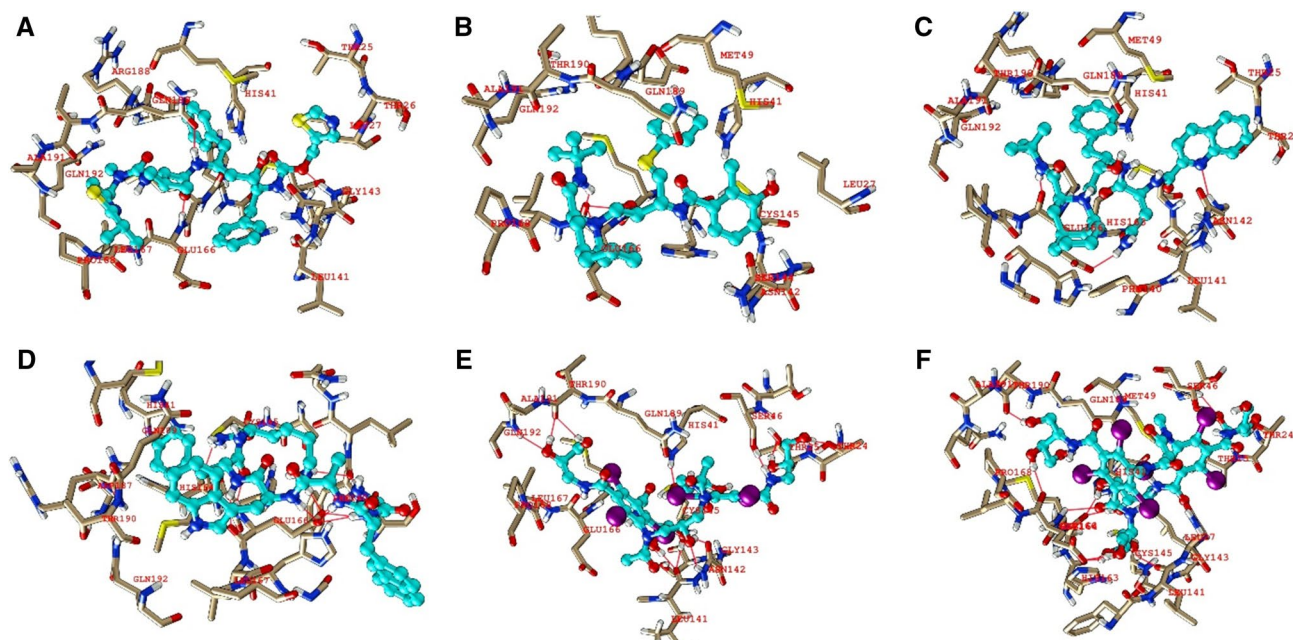
In the case of nelfinavir (**2**), protonated nitrogen of decahydroisoquinoline, NH of *N*-substituted carboxamide and 2-hydroxyl group stabilized the nelfinavir-protein complex by interacting with Glu166 through strong hydrogen bonds. The phenylsulfanyl group of nelfinavir (**2**) was stabilized in the S2 subsite with His41 and Tyr54 amino acids. Additionally, the 3-hydroxy-2-methylphenylformamido group occupied the S1 subsite of the protein active site (Fig. 4B). Similarly in saquinavir (**3**), NH of *tert*-butylcarbamoyl, one of the carbonyl oxygen and NH of butanediamide stabilized saquinavir within the active site of the protein by forming hydrogen bond with Glu166. Quinoline nitrogen of saquinavir further provided stability to the complex by forming a hydrogen bond with Gly143, whereas the phenyl group occupied hydrophobic S2 subsite (Fig. 4C).

Pralmorelin (**4**) also known as growth hormone-releasing peptide 2, used as a diagnostic agent for the assessment of growth hormone deficiency, exhibited promising binding affinity with Mpro protein. The ligand-protein complex was found to be stabilized by nine strong hydrogen

bonds and one salt bridge between pralmorelin and Mpro active site. The hydrogen bonds are mostly with Glu166, Phe140 and His164. The salt bridge was observed between the quaternary terminal nitrogen of pralmorelin and Glu166. Additionally, the phenyl ring occupied the S2 subsite (Fig. 4D).

Apart from these identified anti-viral agents which are already in use as HIV protease inhibitors, some promising interactions of radiocontrast agent iodixanol (**5**) and iotrolan (**6**) were also observed with the target protein under study (Fig. 4E, F, respectively). Both of these are iodine-containing radiocontrast agents injected in the body for visualization in various clinical processes. The iodine-containing drugs carry major limitation for their use, mainly because of adverse effects such as hypersensitivity reactions, cardiovascular, gastrointestinal and neuronal complications (Table 1).

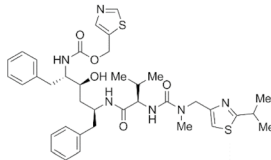
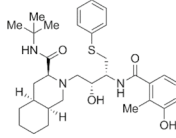
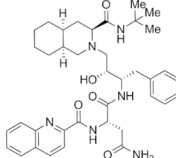
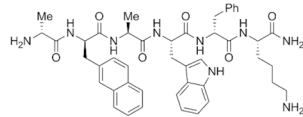
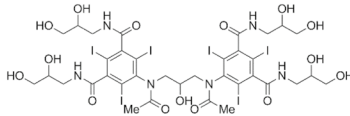
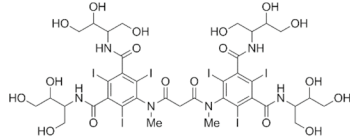
Along with ritonavir, which is already in clinical trials, nelfinavir, saquinavir and pralmorelin can be considered as potential drug candidates for the treatment of COVID-19. Apart from these, the structural modification in iodixanol and iotrolan could lead to stable and safer antiviral agents. Three representative compounds (M1–M3) having modifications in the iodixanol structure are suggested with their respective docking scores in the supporting information. Substitution of the iodo groups in iodixanol with hydrogens could make them stable and safer derivatives, but at the same time, their affinities for the Mpro got compromised a bit (still having promising binding affinities) as shown in Table S2.



**Fig. 4** Docking interactions of compounds (**1–6**: A–F) in the active sites of Mpro. (A) ritonavir, (B) nelfinavir, (C) saquinavir, (D) pralmorelin, (E) iodixanol, (F) iotrolan. Ligands are shown as cyan balls

and sticks. Mpro residues are shown as atom type color sticks. Hydrogen bonds formed between ligands and receptor are indicated by red lines

**Table 1** Chemical structures and docking results of the hits obtained from the screening of approved drug library (approved by FDA or different world authorities)

Compd.	Drug/diagnostic agent	Chemical structure	CAS no.	Glide score (kcal/mol)
1	Ritonavir		155213-67-5	-8.01
2	Nelfinavir		159989-64-7	-7.62
3	Saquinavir		127779-20-8	-7.5
4	Pralmorelin		158861-67-7	-9.17
5	Iodixanol		92339-11-2	-15.77
6	Iotrolan		79770-24-4	-14.36

While repurposing of drugs used in other indications may help to find a faster treatment for COVID-19, it will need a concerted effort to discover a new molecule that specifically targets this virus. Asinex BioDesign library is constructed with 175,815 molecules. This library consists of molecules from natural products which are having pharmacologically important structural features in their chemical scaffolds. To screen this database, the under given protocol was followed. Asinex library of 175,815 compounds was retrieved from the Asinex database and compounds were prepared using LigPrep module for the structural optimization using the OPLS3e force field. Pharmacophore-based screening was performed using this library by Phase module of Schrodinger to obtain Mpro inhibitors with the desired chemical features. A minimum of four-feature matching criteria was set to qualify over the generated seven feature AADRRRR hypothesis. The obtained hits from this screening were ranked orderly by the Phase screen score. From the Asinex library, 26,001 molecules qualified the pharmacophore-based screening criteria. These hits were systematically screened in silico using sequential conformational precision approaches which included HTVS, SP and XP docking of Schrodinger suit on

the Mpro protein, and the molecules obtained in the final step were analyzed manually for their predicted binding modes and scores. Top 20 molecules amongst them are discussed here. All these 20 molecules are divided into four classes based on their chemical structures. All the structures along with their ranking and docking scores (kcal/mol) are mentioned in following section category-wise.

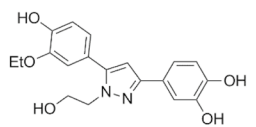
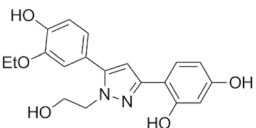
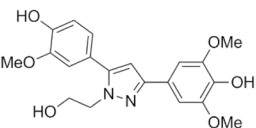
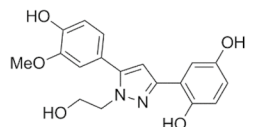
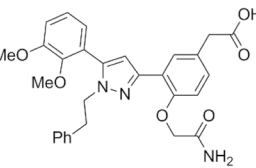
### 3,5-Disubstituted pyrazole-based derivatives (7–11)

The highest-ranked compound (7) was pyrazole-based derivative with substituted phenyl rings at C3 and C5 positions. Analysis of the ligand-receptor complex showed that compound (7) fits snugly into the substrate-binding site (Fig. 5A). The ligand-receptor complex was stabilized by hydrogen bonding between hydroxyl groups of compound (7) and protein. The C3 phenyl ring occupied the S4 subsite of Mpro protein, and the hydroxyl groups at the C3 phenyl ring were found to interact with Thr190 by strong hydrogen bonds. The C5 phenyl ring was observed in alignment with S1 subsite of the protein active site and the hydroxyl group at the C5 phenyl ring imparted stability to the complex by





**Table 2** Chemical structures and docking results of the disubstituted pyrazole-based final hits (**7–11**) obtained from the screening of Asinex BioDesign library

Compd.	Rank	Chemical structure	CAS no.	Glide score (kcal/mol)
<b>7</b>	1		2060132-35-4	−9.02
<b>8</b>	2		2061130-63-8	−8.725
<b>9</b>	4		2061227-92-5	−8.085
<b>10</b>	8		2060714-19-2	−7.575
<b>11</b>	16		2059599-31-2	−7.268

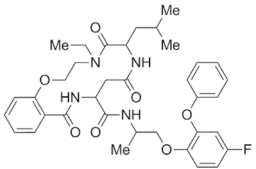
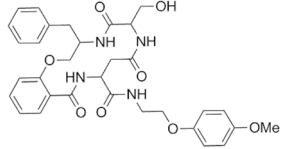
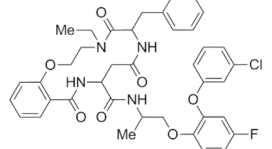
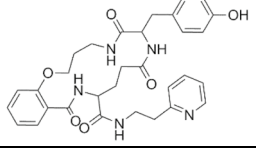
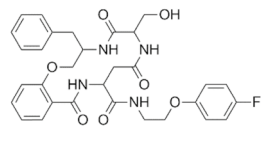
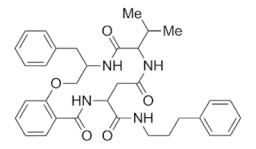
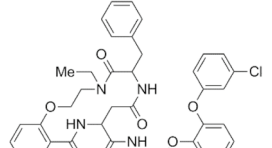
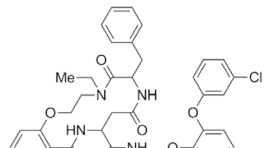
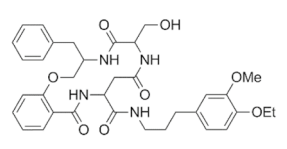
oxygen and Leu141 residue, three NH of the amides and Asn142, Gln189 residues and a stable  $\pi$ - $\pi$  interaction between benzyl ring and His41 residue of the active site. Here, the 4-fluorophenyl ring occupied the S1 subsite, benzyl ring occupied the S2 subsite, and the benzamide moiety was observed in alignment with the S4 subsite of the protein (Fig. 6E). Hydrogen bonding in compound (**17**) was observed between amide nitrogen and Asn142, Gln189 residues and  $\pi$ - $\pi$  interactions between benzyl ring and His41. The subsite occupancy of **17** was similar to that of compound (**16**) (Fig. 6F). Compound (**18**) showed similar interactions as that of the compound (**14**) (Fig. 6G). In this, additional stability was provided by formation of a halogen bond between the chloro group attached to the phenyl ring and NH of Gln192. Compound (**20**) was oriented in a similar fashion to that of compound (**13**), but it showed slightly different interactions (Fig. 6I). Here, hydroxyl group interacted with Hip164 and phenoxy oxygen with Glu166 by hydrogen bonding additionally. The noncyclic amide oxygen also showed hydrogen bond interaction with Cys145. These cyclic amide derivatives displayed significant interactions within the active site of the Mpro enzyme, shaping them potential candidates for further investigation to find out novel anti-SARS-CoV-2 drugs (Table 3).

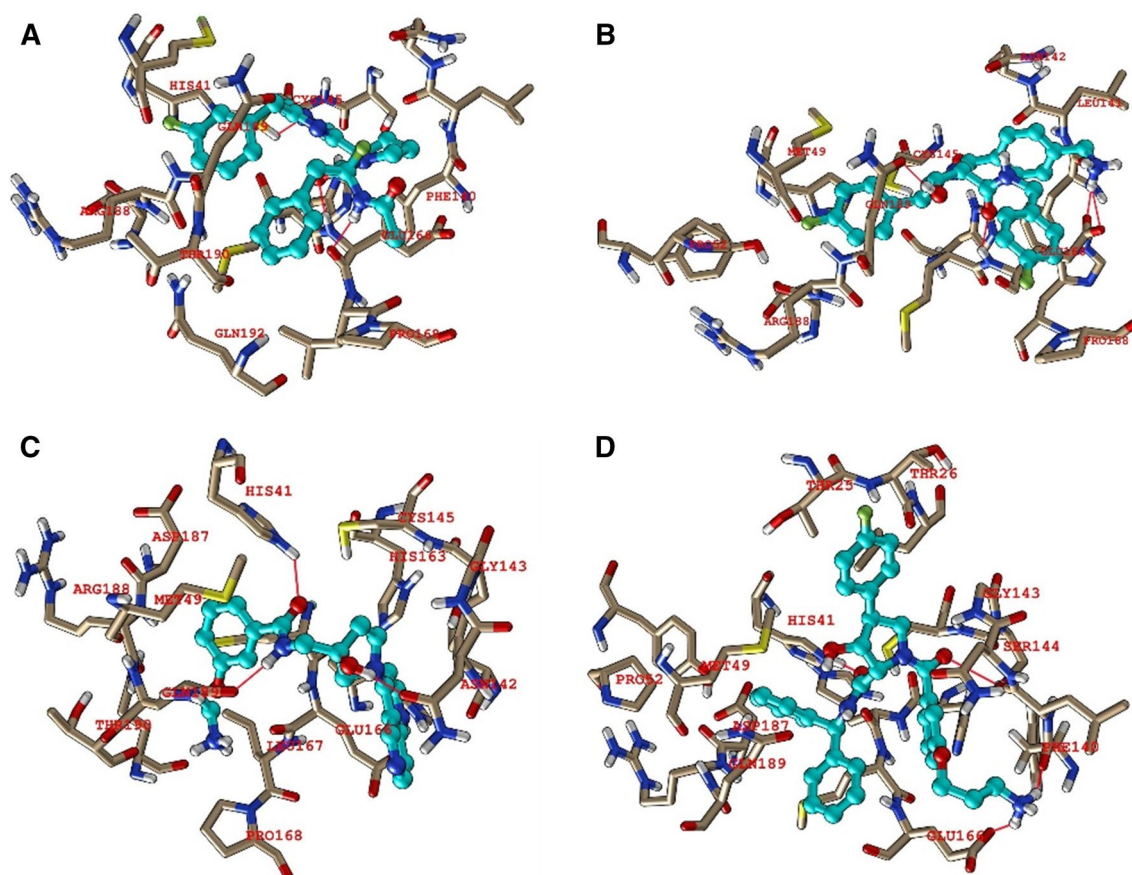
### Pyrrolidine-based derivatives (**21–24**)

Out of the top twenty best scoring compounds, four compounds were observed with pyrrolidine scaffold. Among them, oxygen and NH from two different amide groups of compound (**21**) interacted with Gly166 by hydrogen bonding (Fig. 7A). Further stability to the ligand-receptor complex was provided by hydrogen bond interaction between the nitrogen of pyrazine ring and Cys145. Here, the *m*-fluorobenzyl group occupied hydrophobic S2 subsite, whereas pyrrolidinyl pyrazine scaffold occupied the S1 subsite and benzyl group occupied the S4 subsite for interaction. In compound (**22**), hydrogen bond interaction was observed between the hydroxyl group and Gln189, amide oxygen and Glu166 and a salt bridge between the quaternary amine group and Glu166 (Fig. 7B). Here, *p*-fluorophenylpyrrolidinol scaffold occupied the S2 subsite and the junction of S1-S2 subsite, and *p*-fluorobenzyl amide was observed in alignment with S4 subsite. Likewise, compound (**23**) showed hydrogen bond between the hydroxyl group and Asn142, amide oxygen and His41, amide nitrogen and Gln189. It also exhibited  $\pi$ - $\pi$  interactions between the benzamide ring and His41. Additional stability to this ligand-receptor complex was provided by hydrogen bond and salt bridge interaction between the nitrogen of the naphthyridine ring and Glu166.



**Table 3** Chemical structures and docking results of the cyclic amide based final hits (12–20) obtained from the screening of Asinex BioDesign library

Compd.	Rank	Chemical structure	CAS no.	Glide score (kcal/mol)
12	3		—	−8.273
13	5		—	−7.782
14	9		—	−7.498
15	10		—	−7.467
16	12		—	−7.375
17	17		—	−7.259
18	18		—	−7.228
19	19		—	−7.227
20	20		—	−7.197



**Fig. 7** Docking interactions of compounds (21–24; A–D) in the active sites of Mpro. (A) Compound (21), (B) compound (22), (C) compound (23) and (D) compound (24). Ligands are shown as cyan

balls and sticks. Mpro residues are shown as atom type color sticks. Hydrogen bonds formed between ligands and the receptor are indicated by red lines

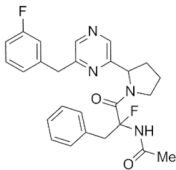
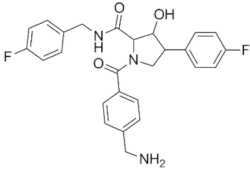
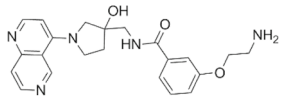
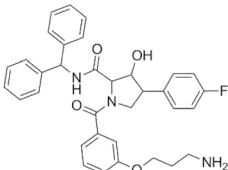
present in the structures, and their interactions have been discussed above. Further to understand the time-dependent stability of the complexes between the first molecule from each discussed group and M<sup>pro</sup> protein structure, molecular dynamics (MD) study was carried out. MD study was performed for a period of 10 ns using Gromacs2020.1 package. Here, considering the docked pose of the ligand-receptor as the reference frame, MD study was performed and various statistical parameters such as RMSD-P, RMSF-P, RMSD-L (*P* = protein, *L* = ligand), H-bonding, Solvent Accessible Surface Area (SASA) etc. were determined.

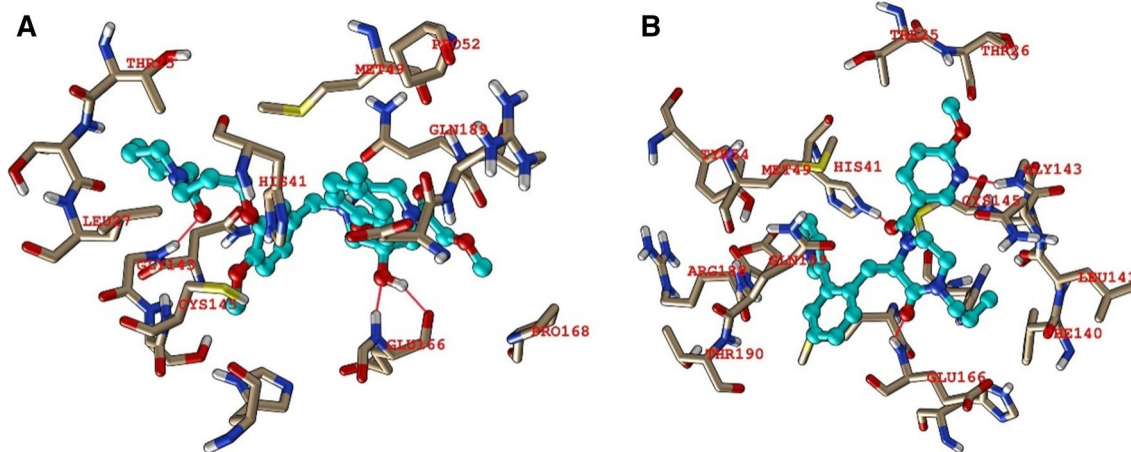
The protein RMSD-P is studied to understand the degree of movement of protein or atoms while putting the ligand in the active site and assessing the structural stability, deviation and conformations of the protein over the simulation time period. The RMSD-P for M<sup>pro</sup> in complexation with compound (7) was in the range of 0.08–0.2 with an average of 0.15 nm (Fig. 9A). This suggests the stability of the protein while having compound (7) in the active site over this period of time. Despite having freely rotatable bonds, the average RMSD-L was 0.35 nm with a range of 0.2 nm to 0.5 nm after

1 ns (Fig. 9B). This strongly suggests the protein–ligand stability without a major change in orientation of the ligand in the active site over the period of simulation time. RMSF explains the structural integrity and residual mobility of the structure. Here again, all the residues including terminal residues and loop regions of the protein structure, while having compound (7) in the active site, showed RMSF-P below 0.41 nm (Fig. 9C).

Hydrogen bonds between protein and compound (7) over the period of analysis were determined with Gromacs *g\_hbond* utility. Maximum seven hydrogen bonds were observed during MD simulation, whereas three to four hydrogen bonds were observed consistently throughout the simulation time (Fig. 9E). Apart from these observations SASA analysis also supported the stability of the protein–ligand complex (Fig. 9D). The short-range electrostatic (Coul-SR) and van der Waals/hydrophobic (LJ-SR) interaction energies between protein and compound (7) explained promising electrostatic as well as hydrophobic interactions. The average values of Coul-SR  $-77.17 \pm 6.7$  kJ/mol and LJ-SR  $-136.45 \pm 4.6$  kJ/mol were observed. This explains

**Table 4** Chemical structures and docking results of the pyrrolidine-based final hits (21–24) obtained from the screening of Asinex BioDesign library

Compd.	Rank	Chemical structure	CAS no.	Glide score (kcal/mol)
21	7		2125619-52-3	-7.676
22	11		2216464-21-8	-7.451
23	14		2060435-10-9	-7.329
24	15		2217261-22-6	-7.288



**Fig. 8** Docking interactions of compounds (25, 26; (A), (B)) in the active sites of Mpro. (A) Compound (25) and (B) compound (26). Ligands are shown as cyan balls and sticks. Mpro residues are shown

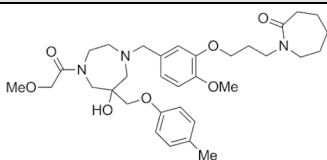
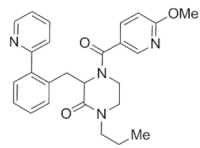
as atom type color sticks. Hydrogen bonds formed between ligands and receptor are indicated by red lines

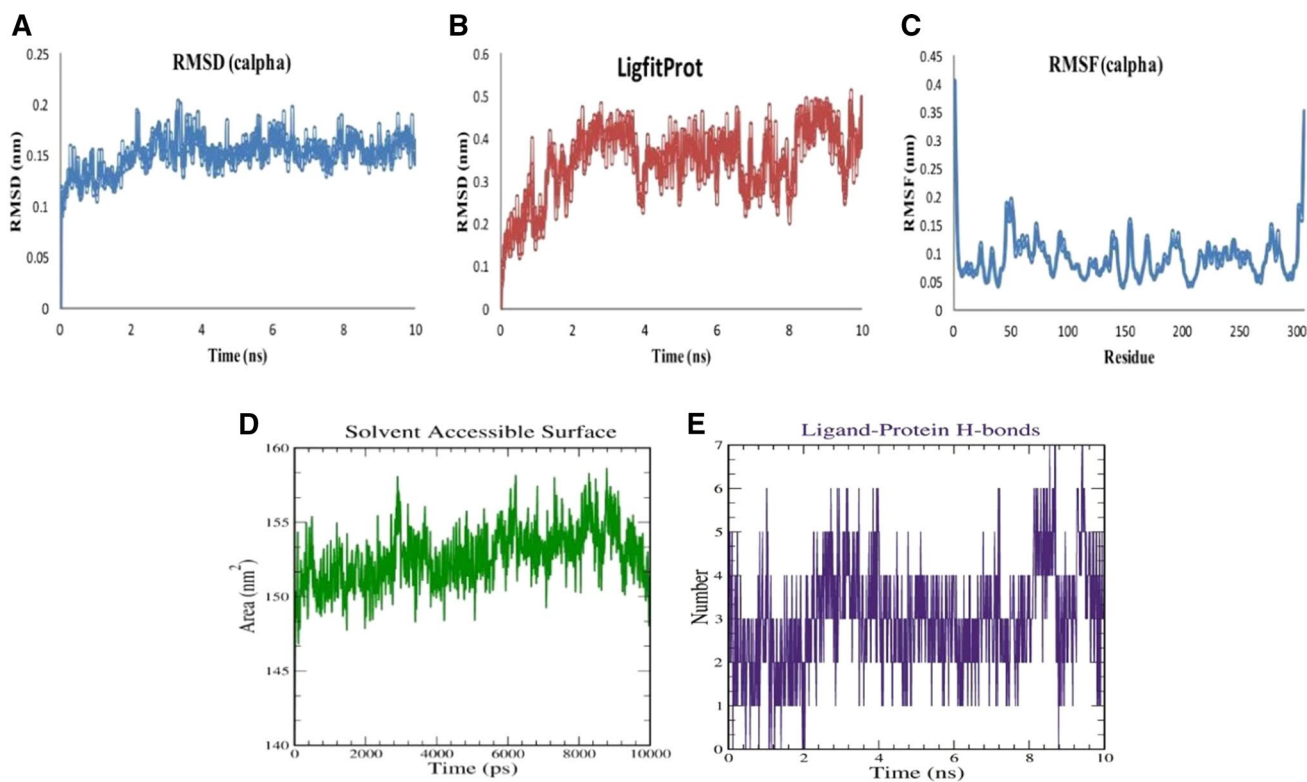
that the role of hydrophobic interaction was more important than the electrostatic interactions in stabilizing the complex.

Similar evaluation was done for Mpro with compound (12). The RMSD-P was observed in the range of 0.1–0.38 nm with an average of 0.24 nm (Fig. 10A). Despite having multiple rotatable bonds and the cyclic amide structure containing a large ring, the RMSD-L for the first ns

was observed in the range of 0.18–0.33, and 2 ns onwards it was consistently observed in the range of 0.26–0.5 nm (Fig. 10B). The observed RMSF-P for the residues up to 300 was below 0.5 nm, whereas the protein tail above residue number 300 showed fluctuation up to 1.5 nm (Fig. 10C). Overall, this ligand-receptor complex showed a maximum of five hydrogen bonds, whereas three to four H-bonds were

**Table 5** Chemical structures and docking results of the miscellaneous final hits (**25**, **26**) obtained from the screening of Asinex BioDesign library

Compd.	Rank	Chemical structure	CAS no.	Glide score (kcal/mol)
<b>25</b>	6		2125474-20-4	-7.763
<b>26</b>	13		—	-7.331

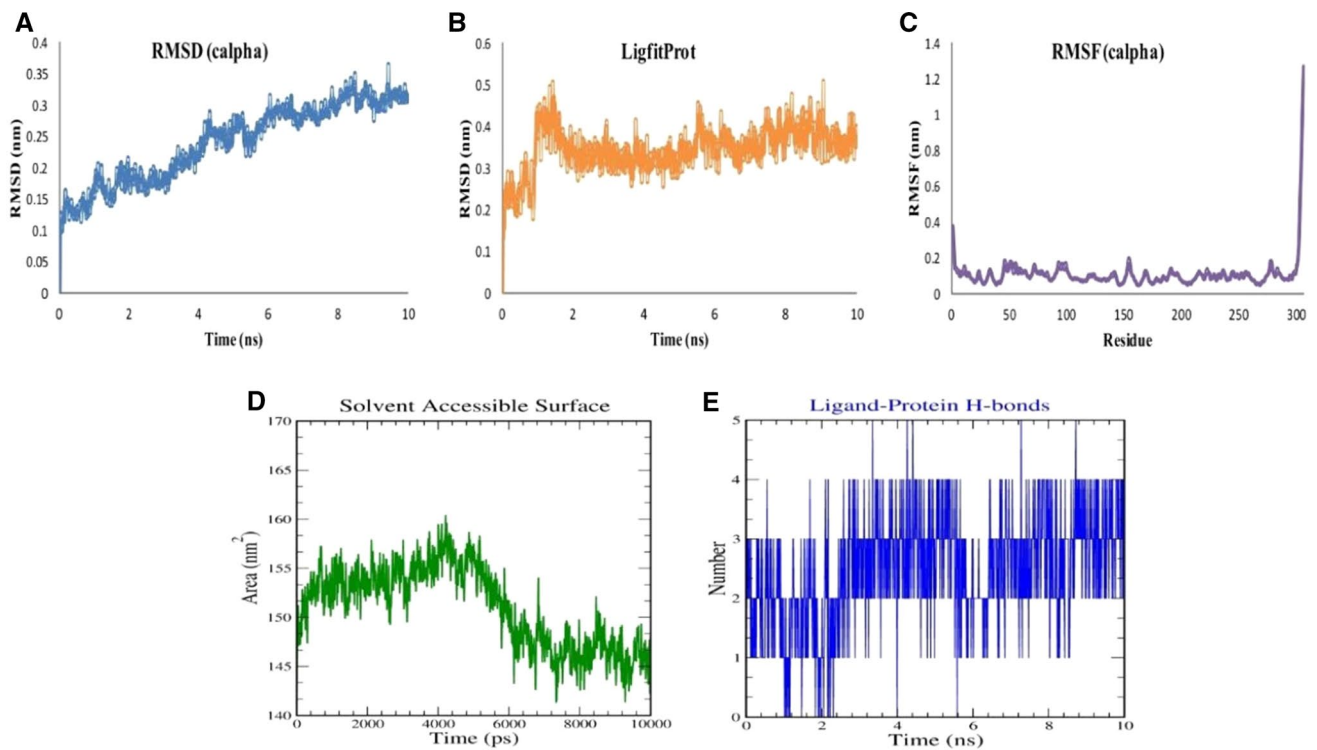


**Fig. 9** RMSD-P, RMSD-L, RMSF-P, SASA and H-bond plots for Mpro with compound (**7**)

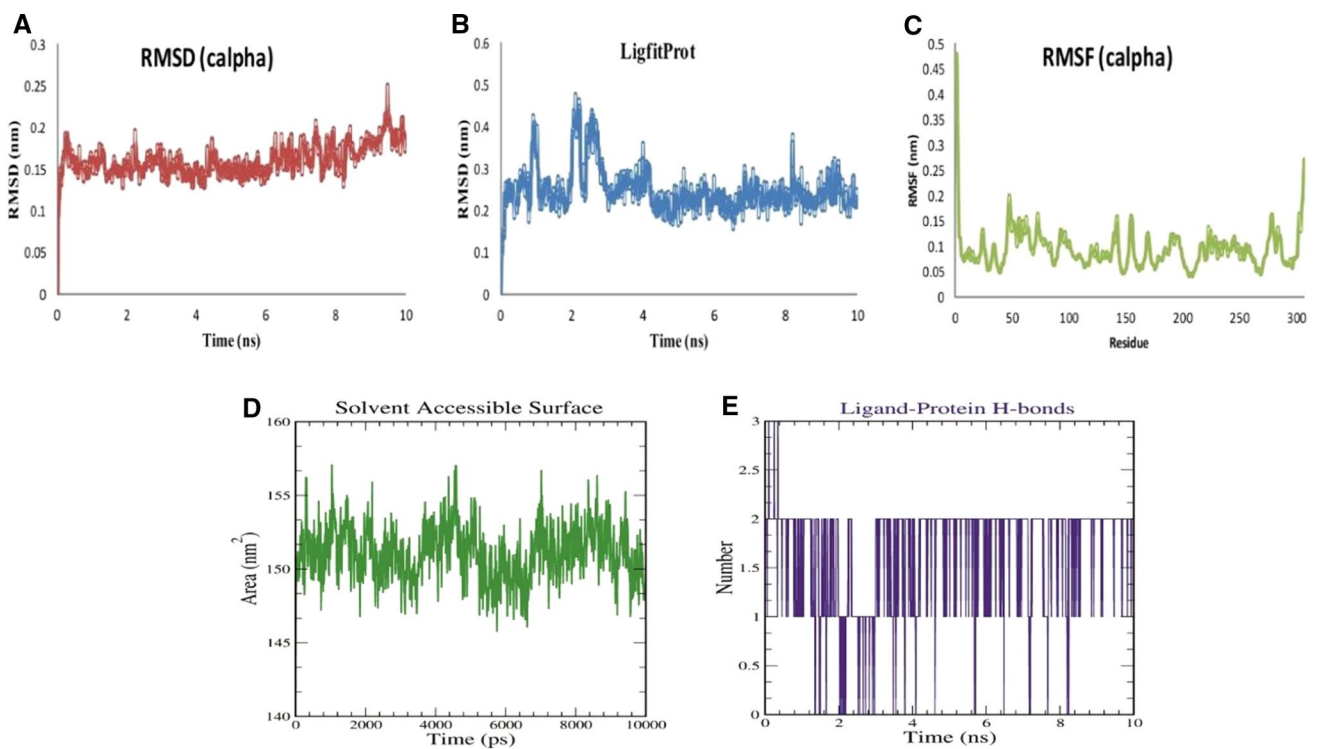
consistently observed over the period of time (Fig. 10E). The short-range electrostatic (Coul-SR, energy:  $-87.66 \pm 5.2$  kJ/mol) and van der Waals/hydrophobic (LJ-SR, energy:  $-179.46 \pm 6.9$  kJ/mol) interaction energies suggested promising interactions between the ligand and the protein. The contribution of hydrophobic interactions was found to be higher than that of the electrostatic interactions.

For compound (**21**) in complexation with Mpro, the RMSD-P values in the range of 0.12–0.25 nm (average 0.16 nm) explained the stability of the protein while having compound (**21**) in the active site (Fig. 11A). Despite high

flexibility of compound (**21**), the RMSD-L was observed in the range of 0.2–0.4 nm consistently except for the time duration from 0.8–1 and 2.2–2.8 ns where a sharp shoot in RMSD-L was observed up to 4.8 nm and it followed normalcy for the remaining period of simulation (Fig. 11B). Here, the overall RMSF-P value below 0.5 nm also supports the stability of the protein in the presence of the ligand in the active site (Fig. 11C). Initial MD analysis showed three hydrogen bonds in the complex, which then stabilized to two for the entire duration of simulation analysis consistently (Fig. 11E). Here again, the short-range electrostatic



**Fig. 10** RMSD-P, RMSD-L, RMSF-P, SASA and H-bond plots for Mpro with compound (12)



**Fig. 11** RMSD-P, RMSD-L, RMSF-P, SASA and H-bond plots for Mpro with compound (21)

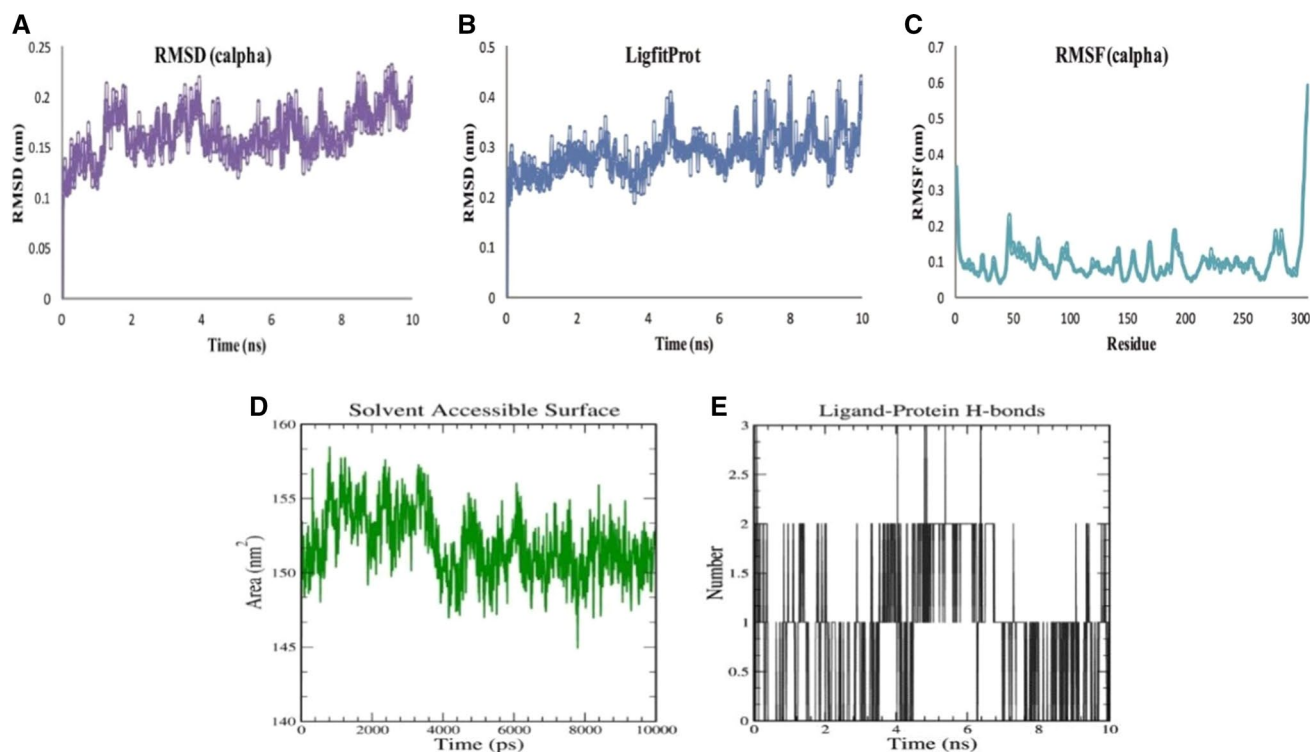
(Coul-SR, energy:  $-74.09 \pm 5$  kJ/mol) and van der Waals/hydrophobic (LJ-SR, energy:  $-146.23 \pm 4.6$  kJ/mol) interaction energies suggested promising interactions between the ligand and the protein. The contribution of hydrophobic interactions was found to be higher than that of the electrostatic.

The RMSD-P value for Mpro-compound (**25**) complex was observed in the range of 0.1–0.24 nm (average 0.16 nm) suggesting the protein stability (Fig. 12A). The RMSD-L was in the range of 0.2–0.44 nm with an average value of 0.29 nm (Fig. 12B). Except for the terminal residues, the RMSF-P was below 0.25 nm and for the terminal residues, it was below 0.6 nm (Fig. 12C). This complex showed a maximum of three hydrogen bonds, whereas one-two hydrogen bonds were consistently present over the simulation period (Fig. 12E). Herein, the contribution of hydrophobic interactions was also observed to be more than the electrostatic interactions (observed values are Coul-SR, energy:  $-78.84 \pm 12$  kJ/mol and LJ-SR, energy:  $-161.74 \pm 7.2$  kJ/mol).

## Conclusion

To conclude, from the approved drug library, using a systematic screening approach, three anti-viral drugs ritonavir, nelfinavir, saquinavir, which are proven HIV protease

inhibitors, were identified as promising drug candidates for the treatment of SARS-CoV-2. Among them, ritonavir is already in phase II/III and IV clinical trials for the treatment of COVID-19. Apart from ritonavir, nelfinavir and saquinavir can also be considered as potential drug candidates against the novel coronavirus. In addition to these, screening of approved drug library also suggested three diagnostic agents pralmorelin, iodixanol and iotrolan, interacting promisingly with Mpro protein. Pralmorelin is used as a diagnostic agent for the assessment of growth hormone deficiency, whereas iodixanol and iotrolan are iodine-containing radiocontrast agents. These three identified diagnostic agents carry many limitations, and thus structural modifications in these three candidates to overcome major limitations could lead to possible potential molecules. From the systematic virtual screening of Asinex BioDesign library, 20 molecules were identified interacting with viral Mpro protein; these are categorized as disubstituted pyrazoles, cyclic amides, pyrrolidine-based and miscellaneous derivatives. Results from Asinex BioDesign screening suggest high potential of these compounds as probable hits for use in the treatment of this devastating viral problem.



**Fig. 12** RMSD-P, RMSD-L, RMSF-P, SASA and H-bond plots for Mpro with compound (**25**)



## Experimental

### Data collection and preparation

3D structure of Mpro was obtained from RCSB Protein Data Bank (PDB code: 6LU7) [24], prepared to ensure structural correctness for hydrogen consistency, bond orders, steric clashes and charges using protein preparation wizard [22] in Schrodinger Suite supported by OPLS3e force field. Thus, the prepared structures were used for pharmacophore model development as well as for receptor grid generation for the docking protocol. The receptor grid was generated considering the position of the co-crystallized ligand N3 in the active site.

The 3-dimensional approved drug and diagnostic agent library (approved by FDA and other world authorities) was obtained from SuperDRUG2 resource for approved drug [19, 20]. A total of 4369 molecules were prepared for computational study at physiological pH condition by using LigPrep module of Schrodinger [25]. Asinex BioDesign library [21], which consists of pharmacologically important structural features from natural products in the chemical scaffolds having synthetic feasibility, was obtained from Asinex database and 175,815 molecules (database updated in early 2020) were prepared using LigPrep module [25].

### Pharmacophore development

Mpro protein structure is deposited in PDB (PDB code: 6LU7) with N3 peptide as the co-crystallized ligand having an  $IC_{50}$  value of 125  $\mu$ M [14]. Being not so promising activity value, we planned to develop active site-based pharmacophore model. To identify the pharmacophoric features required for interaction, the co-crystallized ligand was knocked out and protein active site-based pharmacophore (e-pharmacophore) model was generated [26, 27]. This e-pharmacophore model was developed using Phase module of Schrodinger using the Glide [28] XP scoring for library of inbuilt fragments. These fragments were docked in the receptor grid generated on the active site. Various fragments with different pharmacophoric features were docked, and common features were chosen which satisfied the criteria of position and direction. On the basis of protein amino acids present, excluded volume was calculated to avoid false selection of ligands during screening of the database. The obtained seven feature (AADDRRRR)-based pharmacophore model was used for virtual screening.

### Virtual screening workflow

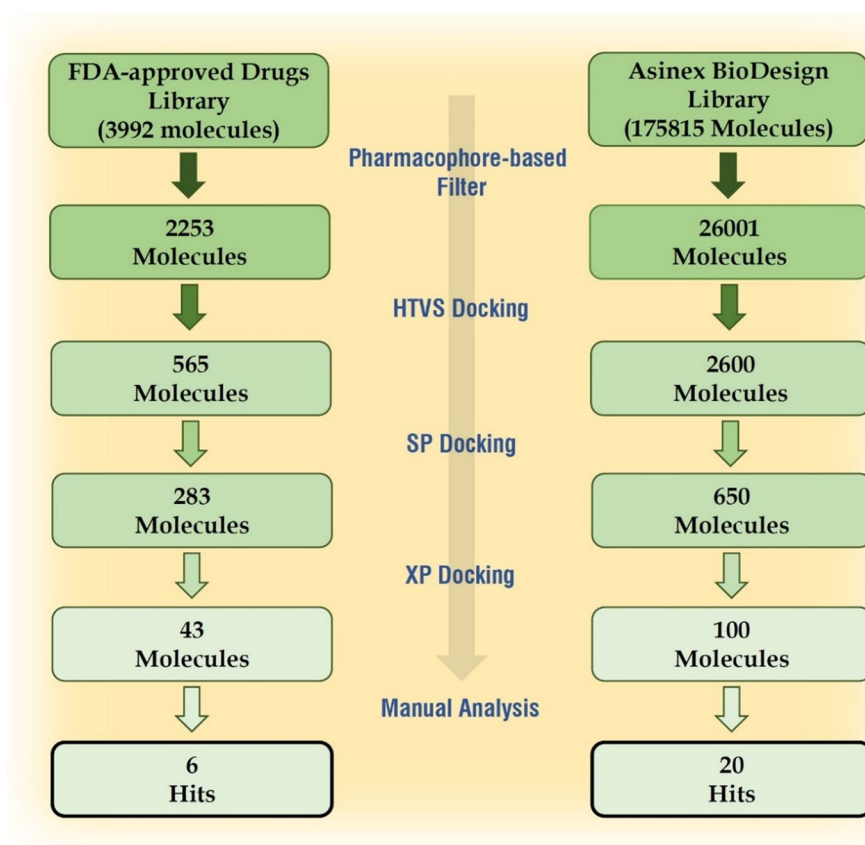
Two selected libraries (FDA-approved drug/diagnostic agents and Asinex BioDesign) were first screened using receptor-based pharmacophore model. A criterion for minimum four-feature match was set, and all the prepared molecules were screened on the obtained AADDRRRR hypothesis. For each molecule, 50 conformers were allowed to be generated and energy minimized. The acceptor and negative as well as donor and positive features were considered equivalent, and scoring function was set to default Phase Screen Score. From the approved drug library, 2253 molecules qualified the pharmacophore-based screening criteria. To them, a considerably fast but raw screening was applied initially by using Glide HTVS and the top 25% molecules were singled out for standard precision (SP) docking in Glide, and from that top 50% were carried forward for extra-precision (XP) analysis. From the XP analysis, top 15% of the compounds were considered further and analyzed manually for their predicted binding modes and scores.

From the Asinex BioDesign library, 26,001 molecules qualified the screen initially, using the pharmacophore-based filter. These compounds were then first passed through HTVS screen and the top 10% molecules were chosen for SP docking, and from these compounds top 25% molecules were moved forward for XP docking studies. Finally, the top 15% molecules were selected for manual analysis of their predicted binding modes and scores (Fig. 13).

### Molecular dynamics

The MD studies between the selected compounds and Mpro (PDB code: 6LU7) were performed for a period of 10 ns by using GROMACS 2020.1 software [29]. The ligand-enzyme complexes as such acquired from the docking studies were considered as the initial points for simulation. To establish the complex stability, CHARMM36 all-atom force field [30] was used to establish the complex stability and ligand-receptor parameters were calculated using GROMACS. CGenFF server was used to generate the ligand topology [31, 32], and the ligand-protein complexes were prepared and solvated using the tip3p/SPC216 water model [33].  $Na^+$  or  $Cl^-$  ions were used to neutralize the system. All the complexes were primarily energy minimized with the steepest descent method [34] followed by two sequential equilibration simulations using canonical (NVT) and isobaric-isothermic (NPT) ensemble for 100 picoseconds (ps) each. Using the NPT group, final MD simulation was carried out and long-range electrostatic interactions were identified by using particle mesh Ewald (PME) method [35]. The MD calculations were carried

**Fig. 13** Schematic representation of virtual screening workflow



out at 300 K temperature and 1 bar pressure using the GROMACS 2020.1 simulation package, and the resulting data were analyzed [36, 37].

**Acknowledgements** Dushyant V. Patel gratefully acknowledges DST-INSPIRE, New Delhi, for awarding Research Fellowship (IF-150660). Nirav R. Patel is thankful to the UGC, New Delhi, for awarding Senior Research Fellowship [F. No. 25-1/2014-15(BSR)/7-129/2007/(BSR)].

**Author contributions** The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### Compliance with ethical standards

**Conflict of interest** All authors declare that they have no conflict of interest.

### References

- Peiris JSM, Guan Y, Yuen KY (2004) Severe acute respiratory syndrome. *Nat Med* 10:S88–S97
- Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367:1814–1820
- Wang C, Horby PW, Hayden FG, Gao GF (2020) A novel coronavirus outbreak of global health concern. *Lancet* 395:470–473
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395:497–506
- Woelfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Mueller MA et al (2020) Clinical presentation and virological assessment of hospitalized cases of coronavirus disease 2019 in a travel-associated transmission cluster. *medRxiv*. 2020.03.05.20030502
- Cucinotta D, Vanelli M (2020) WHO declares COVID-19 a pandemic. *Acta Biomed* 91:157–160
- World Health Organization (2020) Coronavirus disease 2019 (COVID-19) situation report-89. [https://www.who.int/docs/default-source/coronavirus/situation-reports/20200418-sitrep-89-covid19.pdf?sfvrsn=3643dd38\\_2](https://www.who.int/docs/default-source/coronavirus/situation-reports/20200418-sitrep-89-covid19.pdf?sfvrsn=3643dd38_2). Accessed April 19, 2020
- Chen Y, Liu Q, Guo D (2020) Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol* 92:418–423
- Cui J, Li F, Shi ZL (2019) Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 17:181–192
- Fung TS, Liu DX (2014) Coronavirus infection, ER stress, apoptosis and innate immunity. *Front Microbiol* 5:1–13
- Ashour HM, Elkhatib WF, Rahman MM, Elshabrawy HA (2020) Insights into the recent 2019 novel coronavirus (Sars-coV-2) in light of past human coronavirus outbreaks. *Pathogens* 9:1–15
- Thiel V, Ivanov KA, Putics Á, Hertzog T, Schelle B, Bayer S et al (2003) Mechanisms and enzymes involved in SARS coronavirus genome expression. *J Gen Virol* 84:2305–2315
- Zhang L, Lin D, Sun X, Rox K, Hilgenfeld R (2020) X-ray structure of main protease of the novel coronavirus SARS-CoV-2 enables design of  $\alpha$ -ketoamide inhibitors. *bioRxiv* (Internet). 2020.02.17.952879

14. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y et al (2020) Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* 582:286–293
15. Adem S, Eyupoglu V, Sarfraz I, Rasul A, Ali M (2020) Identification of potent COVID-19 main protease (Mpro) inhibitors from natural polyphenols: an in silico strategy unveils a hope against CORONA. Preprints (internet). 2020030333
16. ul Qamar MT, Alqahtani SM, Alamri MA, Chen L-L (2020) Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *J Pharm Anal.* <https://doi.org/10.1016/j.jpha.2020.03.009>
17. Rut W, Groborz K, Zhang L, Sun X, Zmudzinski M, Hilgenfeld R et al (2020) Substrate specificity profiling of SARS-CoV-2 Mpro protease provides basis for anti-COVID-19 drug design. *bioRxiv*. 2020.03.07.981928
18. Liu C, Zhou Q, Li Y, Garner LV, Watkins SP, Carter LJ et al (2020) Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. *ACS Cent Sci* 6:315–331
19. Siramshetty VB, Eckert OA, Gohlke BO, Goede A, Chen Q, Devarakonda P et al (2018) SuperDRUG2: a one stop resource for approved/ marketed drugs. *Nucl Acids Res* 46:D1137–D1143
20. Goede A, Dunkel M, Mester N, Frommel C, Preissner R (2005) Super drug: a conformational drug database. *Bioinformatics* 21:1751–1753
21. Asinex. <http://www.asinex.com>. Accessed on April 05, 2020
22. Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2016; Impact, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY, 2020
23. ClinicalTrials. <https://clinicaltrials.gov>. Accessed April 10, 2020
24. Protein Data Bank. [www.rcsb.org/pdb/home/home.do](http://www.rcsb.org/pdb/home/home.do)
25. LigPrep, Schrödinger, LLC, New York, NY, 2020
26. Loving K, Salam NK, Sherman W (2009) Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *J Comput Aided Mol Des* 23:541–554
27. Salam NK, Nuti R, Sherman W (2009) Novel method for generating structure-based pharmacophores using energetic analysis. *J Chem Inf Model* 49:2356–2368
28. Glide, Schrödinger, LLC, New York, NY, 2020
29. Lindahl, Abraham, Hess, van der Spoel, GROMACS 2020.1 Source code (Version 2020.1). Zenodo. <http://doi.org/10.5281/zenodo.3685919>. Access on March 3, 2020
30. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, Grubmüller H, MacKerell AD (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 14:71–73
31. Vanommeslaeghe K, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr (2009) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM ALL-atom additive biological force fields. *J Comput Chem* 31:671–690
32. Yu W, He X, Vanommeslaeghe K, MacKerell AD (2012) Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J Comput Chem* 33:2451–2468
33. Mark P, Nilsson L (2001) Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J Phys Chem A* 105:9954–9960
34. Bixon M, Lifson S (1967) Potential functions and conformations in cycloalkanes. *Tetrahedron* 2:769–784
35. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092
36. XMgrace. <http://plasma-gate.weizmann.ac.il/Grace/>
37. Rawat R, Verma SM (2020) HeroMDAnalysis: an automagical tool for MD simulation analysis. <http://www.heromdanalysis.wordpress.com>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.